Prospective Indexing for Enhanced Retrieval of Medical Documents on the World Wide Web

Catherine L. Schell, MLS, MD, FAAFP
Richard J. Rathe, MD, FAAFP
Office of Medical Informatics, University of Florida, Gainesville, Florida

Background. Currently quality bibliographic control is dependent on classifications systems like those of National Library of Medicine or the Library of Congress. These systems impose key words on an article after its publication. World Wide Web search engines rely on free text so they are less precise in document retrieval. Web crawlers, spiders, and robots thus tend to be "haphazard." So, "how can physicians access information quickly and reliably..." on the Web? (1) Even if search engines were designed to employ structured vocabulary or "lexical disambiguation," (2) they still would be doing retrospective work. Retrospective indexing by retrofitting key words or concepts to an article after it is published necessitates a time-lag from days to months. With ready access to information provided by the Web, there is a need to eliminate the time-lag in retrospective indexing. Prospective indexing, defined as indexing done prior to publication, may be one solution. With the need to improve bibliographic control of the Internet, (3) there is room for using prospective indexing. Yet we found no designed prospective models.

System. This poster illustrates one model for prospective indexing. NLM's MeSH (medical subject headings) have proven their utility in retrieving medical documents. This model first presents a section of the major MeSH on the Web. Second it relies on voluntary author selection of the Web presented MeSH. When scrolling through the Net, you are your own editor. As a Web author you will be your own preliminary indexer in this model. An incentive for being your own indexer is improving the speed and accuracy by which others on the Internet find your document. Third, the model designs software to assist an author to include controlled MeSH. The software program not only keeps track of chosen MeSH, but also recommends specific major medical headings based on the author's selection. Links to the MeSH thesaurus and MeSH scope notes may be provided with the software's help. After the author has selected the appropriate key words, the program then formats each selected MeSH term in HTML as a meta tag. The software automatically attaches the author selected and

approved MESH terms in HTML format the beginning of a document. The poster will demonstrate the actual HTML mechanism that allows this key word assignment independent of the document. We call this "meta-tagging" MeSH. Since the software formats the author's chosen MeSH automatically in HTML, when the author releases the document to the WWW, it already contains controlled MeSH in HTML format ready for any index robot to capture prospectively. The software could also make helpful suggestions to assist the Web author-indexer. For example, most robots limit the number of characters they acquire in searching to about 1000. The software would notify the author-indexer who reaches the limit. Also the software might be designed to automatically assign the next superior general subject heading level. If the author selects "Liver Diseases, Parasitic" then the program includes "Parasitic Diseases" since current Web search engines cannot "explode" index terms to catch all words under a general heading. The program design could also prompt the author for any relevant standard subheadings such as "Etiology." The software program should be designed to take the Web author-indexer as little time as possible yet provide flexibility for either quick MeSH application or in depth MeSH indexing. Once on the Web, the author's indexing may be scanned for accuracy.

Conclusion. In this model, indexing is done prior to publication. The proposed model presents an attempt at demonstrating a move from retrospective to prospective indexing to enhance medical document retrieval on the WWW.

References
1. Taubes, G. Indexing the Internet. Science, 1995 Sept. 8;269:1354-1356.
2. Peters, R, Sikorski, R. A Guide to Interactive Medicine Navigating to Knowledge. Tools for Finding Information on the Internet. JAMA, 1996 Feb 12;227(6):505-506.
3. Dorman, D. Information Technology. American Libraries, 1997 March; 28(3):78.